Multiple Chromosomes in Bacteria: The Yin and Yang of *trp* Gene Localization in *Rhodobacter sphaeroides* 2.4.1

Chris Mackenzie, Adrian E. Simmons and Samuel Kaplan

Department of Microbiology and Molecular Genetics, University of Texas Medical School, Houston, Texas 77030

Manuscript received March 12, 1999
Accepted for publication June 9, 1999

ABSTRACT

The existence of multiple chromosomes in bacteria has been known for some time. Yet the extent of functional solidarity between different chromosomes remains unknown. To examine this question, we have surveyed the well-described genes of the tryptophan biosynthetic pathway in the multichromosomal photosynthetic eubacterium *Rhodobacter sphaeroides* 2.4.1. The genome of this organism was mutagenized using Tn*5*, and strains that were auxotrophic for tryptophan (Trp$^-$) were isolated. Pulsed-field gel mapping indicated that Tn*5* insertions in both the large (3 Mb CI) and the small (0.9 Mb CII) chromosomes created a Trp$^-$ phenotype. Sequencing the DNA flanking the sites of the Tn*5* insertions indicated that the genes *trpE-yibQ-trpGDC* were at a locus on CI, while genes *trpF-aroR-trpB* were at locus on CII. Unexpectedly, *trpA* was not found downstream of *trpB*. Instead, it was placed on the CI physical map at a locus 1.23 Mb away from *trpE-yibQ-trpGDC*. To relate the context of the *R. sphaeroides trp* genes to those of other bacteria, the DNA regions surrounding the *trp* genes on both chromosomes were sequenced. Of particular significance was the finding that *rpsA1*, which encodes ribosomal protein S1, and *cmkA*, which encodes cytidylate monophosphate kinase, were on CII. These genes are considered essential for translation and chromosome replication, respectively. Southern blotting suggested that the *trp* genes and *rpsA1* exist in single copy within the genome. To date, this topological organization of the *trp* "operon" is unique within a bacterial genome. When taken with the finding that CII encodes essential housekeeping functions, the overall impression is one of close regulatory and functional integration between these chromosomes.

TEN years ago, the first description of a bacterium possessing multiple chromosomes was published (Suwanto and Kaplan 1989a,b). This organism, *Rhodobacter sphaeroides* 2.4.1, was shown to possess two circular chromosomes of 3.0 (CI) and 0.9 (CII) Mb in size. Until that time, the dogma that bacteria always had one circular chromosome went unquestioned.

R. sphaeroides is a photosynthetic member of the α-3 group of Proteobacteria (Woese *et al.* 1990). Other members of this group, *i.e.*, *Agrobacterium tumefaciens* (Allardet-Servent *et al.* 1993), *Brucella melitensis* (Michaux *et al.* 1993), *Paracoccus denitrificans* (Winterstein and Ludwig 1998), and *Ochrobactrum anthropi* (Jumas-Bilak *et al.* 1998) have also been shown to possess multiple chromosomes. However, bacteria possessing multiple chromosomes are not unique to the α-3 group. It has also been shown that *Leptospira interrogans* (Baril *et al.* 1992), *Burkholderia cepacia* (Rodley *et al.* 1995), and more recently *Vibrio cholerae* (Trucksis *et al.* 1998), members of the Spirochaetales β- and γ-Proteobacteria, respectively, also possess multiple chromosomes. Therefore, a decade after the initial discovery, the existence of multiple chromosomes in bacteria is known to be widespread. What remains unclear is the evolutionary selection for this genomic architecture and its biological significance.

We have examined a number of genes of *R. sphaeroides* and have found that some genes occur in multiple copies that are distributed between the two chromosomes. These include three rRNA operons (*rrnA*, *rrnB*, and *rrnC*), each encoding in the following order: 16S rRNA, $tRNA^{Ile}$, $tRNA^{Ala}$, 23S rRNA, 5S rRNA, and $tRNA^{Fmet}$. One of these rRNA operons, *rrnA*, is found on CI, whereas *rrnB* and *rrnC* are found on CII (Dryden and Kaplan 1990). Other genes, including *hemA*/*hemT* (5-aminolevulinic acid synthase, Neidle and Kaplan 1993), *rdxA*/*rdxB* (redox sensors, Neidle and Kaplan 1992), $rpoN_I$/$rpoN_{II}$ (sigma factors), $groEL_I$/$groEL_{II}$ (chaperones) and many genes encoding the enzymes of the reductive Calvin cycle pathway $cbbA_I$/$cbbA_{II}$ (Hallenbeck *et al.* 1990b; Tabita *et al.* 1992), and $cbbP_I$/$cbbP_{II}$ (Hallenbeck *et al.* 1990a; Tabita *et al.* 1992) have also been shown to be duplicated between CI and CII, respectively. The isozymes encoded by these duplicate genes are often structurally similar, but in most cases have been shown to be differentially regulated.

Sequence sampling of CII-specific cosmids has revealed database matches to several hundred known genes (Choudhary *et al.* 1997, 1999). This suggests that CII is in many respects like any other bacterial chromosome. It was also shown that many sequences

Corresponding author: Samuel Kaplan, Department of Microbiology and Molecular Genetics, University of Texas Medical School, 6431 Fannin St., Houston, TX 77030.
E-mail: skaplan@utmmg.med.uth.tmc.edu

present on CII do not return database matches. This suggests that many functions apparently unique to this organism reside on CII. Therefore, CII is not a truncated copy of its larger CI "sib." This was verified when Tn*5* insertions into CII were shown to result in different auxotrophic phenotypes (Choudhary *et al.* 1994). This suggests that essential, nonduplicated housekeeping functions are encoded on this chromosome.

The synthesis of the aromatic amino acids phenylalanine, tyrosine, and tryptophan, and a number of other aromatic compounds, initially share a common biosynthetic pathway that has been studied extensively (Crawford 1989; Nichols 1996; Pittard 1996). The tryptophan branch of the pathway begins by the action of anthranilate synthetase (encoded by *trpD* and *trpE*) on the substrates chorismate and l-glutamine. The subsequent and sequential actions of the products of the *trpG*, *trpF*, and *trpC* genes lead to the penultimate compound in the pathway, indole-3-glycerol phosphate (IGP). In the last step, IGP and l-serine are the substrates for tryptophan synthase. This enzyme is a heterotetramer ($\alpha_2\beta_2$) in which the α and β subunits are encoded by the genes *trpA* and *trpB*, respectively. With the exception of *Acinetobacter calcoaceticus* (Kishan and Hillen 1990), the *trpBA* genes have been shown to be adjacent and usually cotranscribed in that order.

In this article, we demonstrate that the *R. sphaeroides* genes encoding the enzymes of the tryptophan pathway are distributed between the two chromosomes. The genes *trpA* and *trpE-yibQ-trpGDC* are at two distant loci on CI, while *trpF-aroR-trpB* are at a single locus on CII. The genes *trpF* and *trpB* are separated by a hypothetical gene that we have designated *aroR*. In addition to the genes of the tryptophan pathway, we also describe neighboring genes, including *cmkA* and *rpsA1*. In *Escherichia coli* these genes are essential for chromosome replication and translation, respectively. Southern hybridization suggests that there is a single copy of *rpsA1* and it is located on CII, upstream of *trpF-aroR-trpB*. To date, this is a unique genomic arrangement for the *trp* "operon" and the first demonstration in bacteria of genes that encode a single biosynthetic pathway that is distributed between two chromosomes.

MATERIALS AND METHODS

Bacterial strains, cosmids, and plasmids: Those used are listed in Table 1. Unless otherwise stated, the bacterial strains were grown as follows: *R. sphaeroides* 2.4.1 and derivative strains were grown at 30° in either Luria-Bertani (LB) medium or Sistrom's minimal medium A (SMM) supplemented where appropriate with antibiotics: streptomycin/spectinomycin (Sm/Sp) 50 μg/ml, potassium tellurite K_2TeO_3 (Te) 10 μg/ml, tetracycline (Tc) 1 μg/ml, and trimethoprim (Tp) 50 μg/ml. Media for the growth of auxotrophs were supplemented with 20 μg/ml l-tryptophan. *E. coli* strains were grown at 37° in LB medium supplemented where appropriate with antibiotics: ampicillin (Ap) 100 μg/ml, Tc 15 μg/ml, and Sm/Sp 50 μg/ml. *E. coli* strains DH5α and DH5α*phe*$^-$ were used for routine cloning. *E. coli* S17-1 was used for transferring mobilizable plasmids and cosmids to *R. sphaeroides*. Bacterial conjugation was carried out on LB plates (without antibiotics) at 30°.

Isolation of auxotrophs: Bacterial conjugation was carried out as described previously (Donohue and Kaplan 1992; Choudhary *et al.* 1994). The mobilizable suicide plasmid pSUPTn*5*TpMCS that carries the transposon Tn*5*TpMCS (Tn*5*) was introduced into *R. sphaeroides* 2.4.1ΔS by mating from *E. coli* S17-1. Matings were carried out, and exconjugants were plated on LB Te Tp plates. Colonies were replica plated to minimal SMM Te Tp plates. Colonies that were auxotrophic were purified and tested for their ability to grow without tryptophan. This led to the isolation of Trp$^-$ strains CM01, CM02, CM03, CM05, and CM06.

The transposon used has three features relevant to this report: it carries a Tp-resistance (Tp^r) gene; it has a unique *Eco*RI site outside the Tp^r gene; and it has sites for the restriction enzymes *Ase*I, *Dra*I, *Sna*BI, and *Spe*I. These sites occur rarely in the *R. sphaeroides* genome (Suwanto and Kaplan 1989a,b). Digestion of chromosomal DNA from Tn*5* insertion strains (using these enzymes) followed by transverse alternating field electrophoresis (TAFE) permitted the site of Tn*5* insertion to be determined.

Cloning the *R. sphaeroides* regions flanking Tn*5* insertions: The DNA flanking the sites of transposon insertion was cloned as described previously in detail (Mackenzie *et al.* 1995). This method was used to generate plasmids pCM01, pCM02, pCM03, pCM04, pCM05, and pCM06. In the case of plasmid pCM02Sal, the enzyme *Sal*I (there are no sites for this enzyme in the Tn*5*) was used to subclone the intact transposon with flanking *R. sphaeroides* DNA.

Mapping sites of Tn*5* insertion by TAFE gel electrophoresis: DNA plugs were prepared and then digested as described previously (Mackenzie *et al.* 1995). Fragments were resolved on a 1× TBE and 1% SeaKem GTG gel (FMC, Rockland, ME) using a GeneLine II TAFE System. Electrophoresis was carried out at 10° using the following pulse conditions: stage 1, 30-sec pulse for 5 hr at 350 mA; stage 2, 45-sec pulse for 8 hr at 370 mA; stage 3, 60-sec pulse for 8 hr at 370 mA; stage 4, 90-sec pulse for 5 hr at 390 mA.

DNA sequencing of complete genes and chromosomal regions: Plasmids pCM01, pCM02, pCM03, pCM04, pCM05, and pCM02Sal, as well as cosmids pUI8668 and pLX1P20, were used for further subcloning of the region surrounding *trpEGDC* for sequencing. Cosmid pUI8063 and pUI8536 were used in the same way for regions surrounding *trpA* and *trpFB*, respectively. Successive rounds of cloning and sequencing allowed the appropriate DNA fragments to be sequenced on both strands.

Sequencing reactions: Plasmid DNA was prepared using Wizard Plus SV minipreps (Promega, Madison, WI). Sequencing of Tn*5*-*R.sphaeroides* hybrid fragments used three primers, GW25 (5′-TTCAGGACGCTACTTGTGTA-3′), which is complementary to the IS*50* of the transposon, and pBluescript T3 and T7 primers. GW25 was used to sequence from the transposon into the flanking *R. sphaeroides* DNA. All other plasmid-sequencing reactions used the T3 and T7 primers alone. PCR products were sequenced using specific PCR primers (Table 2). DNA sequencing was performed at the Microbiology and Molecular Genetics Core Facility using Big-Dye chemistry and an ABI 377A sequencer (Applied Biosystems, Foster City, CA).

PCR: Reactions had the following components: 200 μm dNTPs, 25 pmol of each primer (Table 2), 5% v/v DMSO, 10 ng of template DNA, and 2.5 units of *Pfu* DNA polymerase (Stratagene, La Jolla, CA). Cycling times were as follows: step 1, 95°, 1 min; step 2, 15° below lowest primer Tm, 1 min; step 3, 72°, 2 min; step 4, 24 times to step 1; step 5, 72° for 4 min

TABLE 1

Strains

Strain, cosmid, or plasmid	Relevant genotype (phenotype)	Source[a]
R. sphaeroides strains		
2.4.1	Wild type	
2.4.1ΔS	2.4.1 lacking 42-kb plasmid	Suwanto and Kaplan (1992)
CM01	2.4.1ΔS CI *trpE*::Tn*5*TpMCS (Tp^r Trp^-)	This work
CM02	2.4.1ΔS CI *trpD*::Tn*5*TpMCS (Tp^r Trp^-)	This work
CM03	2.4.1ΔS CI *ybaU*::Tn*5*TpMCS (Tp^r Trp^-)	This work
CM04[b]	2.4.1ΔS CI *argDF*::Tn*5*TpMCS (Tp^r Arg^-)	This work
CM05	2.4.1ΔS CI *ybaU*::Tn*5*TpMCS (Tp^r Trp^-)	This work
CM06	2.4.1ΔS CII *trpB*::Tn*5*TpMCS (Tp^r Trp^-)	This work
CM07	CM06 complemented with cosmid pUI8536 (Tp^r Tc^r Trp^+)	This work
CM08	2.4.1 *aroR*::Ω Sm/Sp (Sm^r Sp^r Trp^+)	This work
CM09	2.4.1 CI *trpA*::Ω Sm/Sp (Sm^r Sp^r Trp^-)	This work
CM10	CM09 complemented with cosmid pUI8063 (Tc^r Sm^r/Sp^r Trp^+)	This work
CM11	2.4.1 *trpE*::Ω Sm/Sp (Sm^r Sp^r Trp^-)	This work
CM12	2.4.1 *trpG*::Ω Sm/Sp (Sm^r Sp^r Trp^-)	This work
E. coli strains		
S17-1	*thi pro hsdR hsdM*$^+$ *recA* integrated plasmid RP4-Tc::Mu-Km::Tn*7*	Simon *et al.* (1983)
DH5α	F^-, ϕ80Δ*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA endA hsdR supE44 thi gyrA relA*	
DH5α-Phe^-	F^-, ϕ80Δ*lacZ*ΔM15 Δ(*lacZYA-argF*)U169 *recA endA hsdR supE44 thi gyrA relA phe*::Tn*10d*Cm, Cm^r	J. Eraso (University of Texas Medical School)
XL-Blue MRA	Δ(*mcrA*)183 Δ(*mcrCB-hsdSMR-mrr*)173 *endA1 supE44 thi-1 gyrA96 relA1 lac*	Stratagene
Cosmid and plasmid vectors		
pLA2917	Tc^r	Allen and Hanson (1985)
pLAFRx	Derivative of pLAFR *ori* RK2 (Mob^+ Tc^r)	Heath *et al.* (1992)
pSUPTn*5*TpMCS	Tn5 derivative (Tp^r, unique *Eco*RI site and sites for *Ase*I, *Dra*I, *Sna*BI, and *Spe*I)	Choudhary *et al.* (1994)
pBluescript SK (+) (pBS)	T3 and T7 promoters (Ap^r)	Stratagene
pBluescript II KS (+) (pBSII)	T3 and T7 promoters (Ap^r)	Stratagene
pUI1638	Source of 2082-bp Ω Sm/Sp, see GenBank accession number M60473 (Ap^r Sm^r Sp^r)	
pSUP203	pBR325 derivative (Mob^+ Ap^r Cm^r Tc^r)	
Constructs		
pUI8063	pLA2917 derivative containing CI *trpA* region (Tc^r)	This work
pUI8536	pLA2917 derivative containing CII *trpFB* region (Tc^r)	This work
pUI8668	pLA2917 derivative containing CI region *helO–trpEGD* region (Tc^r)	This work
pLX1P20	pLAFRx derivative containing CI *trpEGDC* (Tc^r)	This work
pCM01[c]	pBSII containing a 7.1-kb *Rs*-Tn*5 Eco*RI fragment from CM01 (Tp^r Ap^r)	This work
pCM02[c]	pBSII containing a 17.4-kb *Rs*-Tn*5 Eco*RI fragment from CM02 (Tp^r Ap^r)	This work
pCM02Sal	pBSII containing a 24-kb *Sal*I fragment from CM02 (Tp^r Ap^r)	This work
pCM03[c]	pBSII containing a 10-kb *Rs*-Tn*5 Eco*RI fragment from CM03 (Tp^r Ap^r)	This work
pCM04[c]	pBS containing an 11-kb *Rs*-Tn*5 Eco*RI fragment from CM04 (Tp^r Ap^r)	This work

(*Continued*)

in a PTC-100 thermal cycler (M.J. Research, Watertown, MA). PCR primer pairs were designed so that the PCR products were from the internal regions of the genes described. The primer pairs are listed in Table 2.

Cosmid libraries: A previously described and ordered CII-specific pLA2917 cosmid library was available (Choudhary *et al.* 1994). A second library was constructed using the cosmid vector pLAFRx, as described by Sambrook *et al.* (1989). After

TABLE 1

(Continued)

Strain, cosmid, or plasmid	Relevant genotype (phenotype)	Source[a]
pCM05[c]	pBSII containing a 16.5-kb *Rs*-Tn*5* *Eco*RI fragment from CM05 (Tp^r Ap^r)	This work
pCM06[c]	pBS containing a 7.0-kb *Rs*-Tn*5* *Eco*RI fragment from CM06 (Tp^r Ap^r)	This work
pCM07A	pBS containing a 6.5-kb *Bam*HI fragment containing *trpA* from cosmid pUI8063 (Ap^r)	This work
pCM08A	Insertion of a 1.3-kb *Sma*I fragment from pCM07A containing *trpA* at *Eco*RV site of pBS (Ap^r)	This work
pCM09A	pCM08A containing Ω Sm/Sp inserted into filled *Sty*I site within *trpA* (Ap^r Sm^r Sp^r)	This work
pCM10A	*Pvu*II fragment *trpA*::Ω Sm/Sp from pCM09A inserted into *Sca*I site of pSUP203 (Mob^+ Tc^r Sm^r Sp^r)	This work
pCM11E	pBSII KS containing a 2.9-kb *Pst*I fragment containing *trpE* from pCM03	This work
pCM12E	1.4-kb *Hin*dIII/*Hin*cII fragment (containing *trpE*) from pCM11E inserted into pBSII KS (Ap^r)	This work
pCM13E	pCM12E with Ω Sm/Sp inserted into *Bst*EII site in *trpE* (Ap^r Sm^r Sp^r)	This work
pCM14E	*Pvu*II fragment containing *trpE*::Ω Sm/Sp from pCM13E inserted into *Sca*I site of pSUP203 (Mob^+ Tc^r Sm^r Sp^r)	This work
pCM15G	2.3-kb *Pst*I fragment from pCM03 inserted into *Pst*I site of pBSII	This work
pCM16G	pCM15G with Ω Sm/Sp inserted into *Nsi*I site within *trpG* (Ap^r Sm^r Sp^r)	This work
pCM17G	*Pvu*II fragment *trpG*::Ω Sm/Sp from pCM16G inserted into *Sca*I site of pSUP203 (Mob^+ Tc^r Sm^r Sp^r)	This work
pCM18R	1.3-kb *Bgl*II fragment containing *aroR* from CII cosmid pUI8536 inserted into *Bam*HI site of pBS (Ap^r)	This work
pCM19R	pCM18R with Ω Sm/Sp inserted into *Sma*I site within *aroR* (Ap^r Sm^r Sp^r)	This work
pCM20R	pCM19R *Pvu*II fragment *aroR*::Ω Sm/Sp inserted into *Sca*I site of pSUP203 (Mob^+ Tc^r Sm^r Sp^r)	This work

[a] Unless otherwise stated, strains shown were generated in this study.
[b] Tn5TpMCS insertion is located between *argD* and *argF*.
[c] *Rs*-Tn*5* indicates a *R. sphaeroides*-Tn*5* hybrid fragment.

in vitro packaging with Gigapack Gold III (Stratagene, La Jolla, CA), the cosmids were introduced into *E. coli* XL1-Blue MRA. The cosmids were pronged onto plates, grown overnight, and then screened for colony hybridization as described.

Complementation: Recombinant cosmids pUI8063 and pUI8536 and the vector pLA2917 were introduced into *E. coli* S17-1. They were then mated to an *R. sphaeroides trpA* mutation (CM09 × pUI8063; CM09 × pLA2917) and *trpB*$^-$ mutation (CM06 × pUI8536; CM06 × pLA2917) on LB plates as described previously (Choudhary *et al.* 1994). The *R. sphaeroides* strains were selected initially on LB Tc Te plates. Twenty Tc^r colonies (from four independent crosses) were then tested

TABLE 2

Internal PCR primers

Gene	Left primer	Right primer
trpA	5′-GTCGCCTATATCATGGCGGG-3′	5′-AGATGCCGAAGCCGACGAT-3′
trpB	5′-GAACATGCCAAGACCGACCC-3′	5′-GCGCGATCTTGATGACATGG-3′
trpC	5′-CTGCCGCATCAGGCTCTC-3′	5′-ATCCTGTCCGGTGCGTCC-3′
trpD	5′-ACAGGATGCTGTCGGTGG-3′	5′-GGCTGAAGCCCCTGATCG-3′
trpE	5′-TTGACCGTCTCCTGATATTCG-3′	5′-GGGCAGAACCAGATCGTCTA-3′
trpF	5′-GCAGGAGTCCGCGTCAAGAT-3′	5′-CATCCAGGGCCTGATCCAGT-3′
trpG	5′-GATGGGCAGGCTCTTGTG-3′	5′-TGCTGCTTCTCATCGACAAC-3′
aroR	5′-ACCTATCTCGCGGCCTTCG-3′	5′-TCAGTCCAGCGCCCGTTTAT-3′
rpsA1	5′-TCGACCTGAAGGAATTCGCC-3′	5′-TGCTCGAGACCCACGAACAG-3′

TABLE 3

Sites of Tn*5* or Ω Sm/Sp insertion

Strain	Gene[a]	Codons[b]	Site of insertion within gene[c]
CM01	*trpE*	503	Tn*5* insertion between codons 253 P and 254 S
CM02	*trpD*	338	Tn*5* disruption of codon 321 S
CM03	*ybaU*	621	Tn*5* disruption of codon 6 S
CM05	*ybaU*	621	Tn*5* insertion between codons 238 T and 239 Y
CM06	*trpB*	409	Tn*5* disruption of codon 91 L
CM08	*aroR*	148	Ω Sm/Sp disruption of codon 47 P
CM09	*trpA*	263	Ω Sm/Sp disruption of codon 164 P
CM11	*trpE*	503	Ω Sm/Sp insertion between codons 171 E and 172 V
CM12	*trpG*	195	Ω Sm/Sp disruption of codon 43 H

[a] The gene containing the insertion.
[b] The number of codons, excluding the termination codon, which comprise the gene.
[c] The position of the Tn*5* or Ω Sm/Sp cartridge insertion within the gene by codon location. The letter that follows the number indicates the amino acid residue(s) disrupted by the insertion (single-letter code).

for their ability to grow on SMM, SMM Tc, and SMM Tc Sm/Sp (Ω::*trpA*$^-$) or SMM Tc Tp (Tn*5*::*trpB*$^-$) plates. Complemented strains could grow on all three plates, *i.e.*, CM10 and CM07.

Southern blotting: Gels were depurinated and then transferred by alkali to Hybond N$^+$ membranes (Amersham, Piscataway, NJ) using standard techniques (Sambrook *et al.* 1989). Probes were radiolabeled using [α-^{32}P]dCTP and a RadPrime DNA labeling system (GIBCO-BRL, Gaithersburg, MD). Probes were purified using a Sephadex G50 spin column and then denatured before use.

Hybridization: Southern blots and colony lifts were hybridized using standard techniques (Sambrook *et al.* 1989). Hybridization and washing were carried out at 55° or 65° for blots and lifts, respectively. Washing was carried out at the hybridization temperature in the following solutions: 4× SSC and 0.1% SDS for 10 min done twice, 1× SSC and 0.1% SDS for 10 min done twice. For lifts, an additional wash in 0.1× SSC and 0.1% SDS for 10 min was done twice.

Cloning *trpA*: Primers were made to the *R. capsulatus trpA* sequence (Table 2), and PCR was performed as described above. The PCR product was gel purified and then sequenced. The DNA sequence matched other *trpA* genes. The PCR product was then used to probe the pLA2917 genomic library. Positively hybridizing cosmids were isolated and probed by Southern hybridization. A positively hybridizing 6.5-kb *Bam*HI fragment was subcloned from cosmid pUI8063 into the *Bam*HI site of pBS to give plasmid pCM07A. A 1.3-kb *Sma*I fragment containing *trpA* from pCM07A was subcloned into the *Eco*RV site of pBS to give pCM08A.

Gene disruptions: The following strategy was used to disrupt the genes *trpA*, *trpE*, *trpG*, and *aroR* (Table 3). An omega (Ω) Sm/Sp cartridge carrying transcriptional terminators was used to make polar gene disruptions. The Ω cartridge, carried on a *Sma*I fragment, was cloned into the gene of interest. The disrupted gene was then subcloned into the mobilizable suicide vector pSUP203. This construct was mated into *R. spheroides* 2.4.1. Exconjugants were selected on LB Sm/Sp tryptophan plates. They were then tested for their ability to grow without tryptophan.

***trpA*:** Plasmid pCM08A was linearized using *Sty*I, which cuts within *trpA* at codon 164. The ends of the linearized plasmid were filled using Klenow fragment and then the Ω Sm/Sp cartridge inserted, giving plasmid pCM09A. This plasmid was digested with *Pvu*II, and a 3.8-kb fragment containing the interrupted *trpA* was inserted into *Sca*I-digested pSUP203 to give pCM10A. This plasmid was introduced into *R. sphaeroides* from *E. coli* S17-1, resulting in *R. sphaeroides trpA* mutation strain CM09.

***trpE*:** A 2.9-kb *Pst*I fragment containing *trpE* from pCM03 was cloned into the *Pst*I site of pBSII to give plasmid pCM11E. A 1.4-kb *Hin*dIII/*Hin*cII fragment containing *trpE* was excised from this plasmid and subcloned into the *Hin*dIII/*Hin*cII sites of pBSII to give pCM12E. An Ω Sm/Sp cartridge was inserted into the *Bst*EII site in *trpE* (between codons 171 and 172) to give plasmid pCM13E. A *Pvu*II fragment from pCM13E was excised and inserted into the *Sca*I site of pSUP203, resulting in plasmid pCM14E. This plasmid was introduced into *R. sphaeroides* from *E. coli* S17-1, resulting in *R. sphaeroides* strain CM11.

***trpG*:** A 2.3-kb *Pst*I fragment containing *trpG* from pCM03 was cloned into the *Pst*I site of pBSII to give plasmid pCM15G. An Ω Sm/Sp cartridge was inserted into the *Nsi*I site within *trpG* (disruption of codon 43) to give plasmid pCM16G. A *Pvu*II fragment containing the disrupted *trpG* was then inserted into the *Sca*I site of pSUP203 to give plasmid pCM17G. This plasmid was introduced into *R. sphaeroides* from *E. coli* S17-1, resulting in *R. sphaeroides* strain CM12.

***aroR*:** Cosmid pUI8536 was digested with *Bgl*II. A 1.3-kb fragment containing *aroR* was inserted into the *Bam*HI site of pBS to give pCM18R. This plasmid was partially digested with *Sma*I. The Ω Sm/Sp cartridge was then inserted. A plasmid containing the Ω insertion at *aroR* codon 47 was selected (pCM19R). This was digested with *Pvu*II, and a 3.8-kb fragment was taken and cloned into the *Sca*I site of pSUP203 to give plasmid pCM20R. This plasmid was introduced into *R. sphaeroides* from *E. coli* S17-1, resulting in *R. sphaeroides aroR* mutation strain CM08.

Sequence analysis: DNA editing was carried out using Seqed (Applied Biosystems). Fragments were then assembled using Gelassemble (version 9.1, Genetics Computer Group, Madison, WI). PCR primers were designed using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi). BLASTX and BLASTP were used for database comparison through the National Center for Biotechnology Information website (http://www.ncbi.nlm.nih.gov/).

Nucleotide accession numbers: The sequences described in results have the following GenBank accession numbers: AF10704, *trpA*; AF107095, *guaB*, *lctD*, *mosC*; AF107096, hypothetical GTP-binding protein; AF108766, *asmA* [partial coding sequence (cds)], *ybaU*, *trpE*, *yibQ*, *trpG*, *trpD*, *trpC*, *moaC*, *lexA*, *comE*, *gluS*, *cisY* (partial cds); AF107093, *cmkA*, *rpsA1*, *hipB*, *trpF*, *aroR*, *trpB*, *Synechocystis* orf.

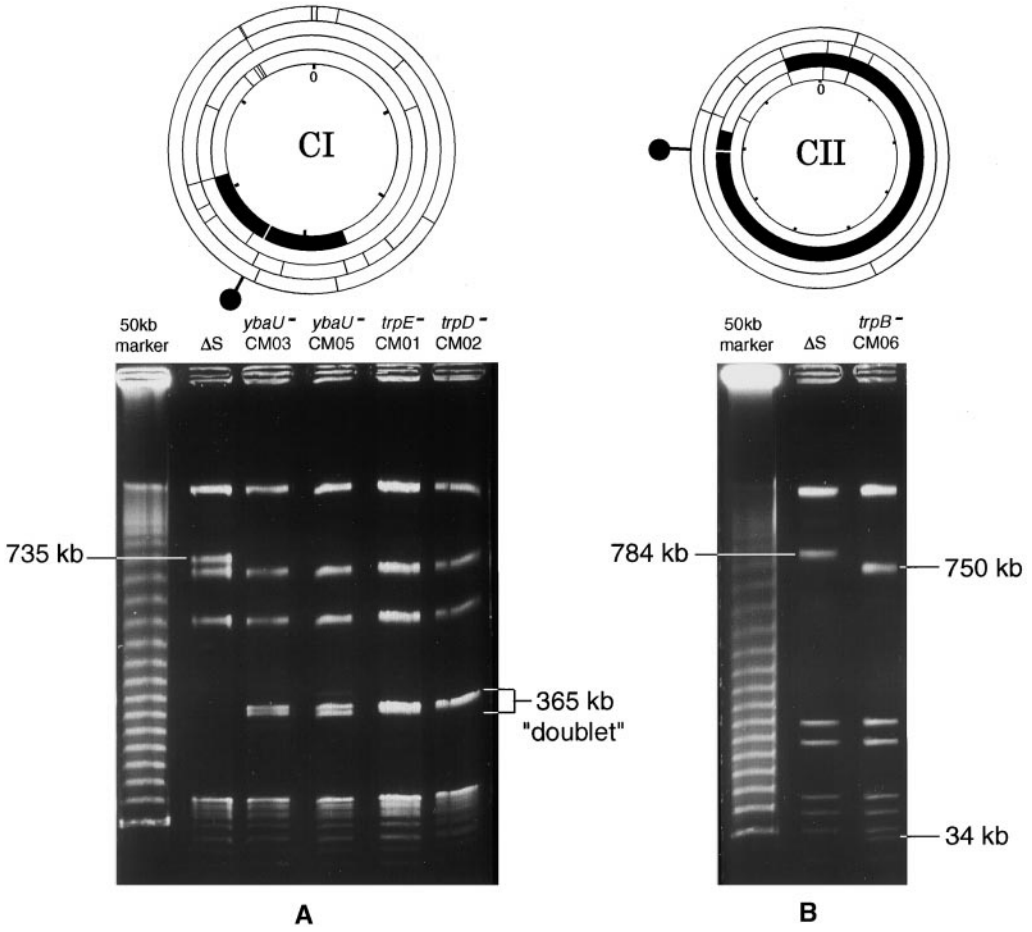

Figure 1.—Mapping the sites of Tn5 insertion. Tn5 insertion strains were made in a 2.4.1ΔS background (see materials and methods). Both panels show a TAFE gel containing DNA from *R. sphaeroides* 2.4.1ΔS and Tn*5* insertion strains that are Trp⁻. The leftmost lane in each gel contains a 50-kb marker ladder of concatenated λ genomes. The lowest rung of the ladder is 50 kb. Above each gel panel is the physical map of CI (left) and CII (right). Each physical map is composed of four concentric circles marked with restriction sites (moving from outer to inner concentric) for enzymes *Ase*I, *Dra*I, *Sna*BI, and *Spe*I. The zero position (0) is shown at 12 o'clock within each map. The distance markers shown around the inner circle are in increments of 500 kb (CI) and 100 kb (CII). The black arc represents the restriction fragment that is being examined in the gel below. The full arc is what is found in 2.4.1ΔS. The two smaller arcs (resulting from cleavage of the full arc due to Tn*5* insertion) represent the new DNA fragments found in the Trp⁻ strains. The site of Tn*5* insertion is shown as a lollipop. It has been placed in this position by using other enzymes for digestion (not shown). (A) *R. sphaeroides* 2.4.1ΔS and CI Tn*5* insertion strains CM03 (*ybaU*⁻), CM05 (*ybaU*⁻), CM01 (*trpE*⁻), and CM02 (*trpD*⁻) are shown. The DNA has been digested with *Spe*I. In 2.4.1ΔS, an *Spe*I band of 735 kb is visible. In the Trp⁻ strains, this band is absent (indicating its disruption by Tn*5*); instead, there is a 365-kb "doublet" that is not visible in the 2.4.1ΔS strain. This indicates that the Tn*5* insertion lies within the 735-kb CI *Spe*I fragment in these Trp⁻ strains. It can be seen that the doublet bands get closer together. This indicates that the Tn*5* insertions in *YbaU* are farther out from the center of the 735-kb *Spe*I fragment than those in *trpE* and *trpD*. These differences in fragment size, though visible, were not sufficient to be estimated on this gel. (B) *R. sphaeroides* 2.4.1ΔS and CII Tn*5* insertion strain CM06 (*trpB*⁻) is shown. The DNA has been digested with *Sna*BI. In 2.4.1ΔS, a *Sna*BI band of 784 kb is visible. In the Trp⁻ strain, this band is absent (indicating its disruption by Tn*5*); instead, there are two bands of 750 and 34 kb that are not visible in the 2.4.1ΔS strain. This indicates that the Tn*5* insertion lies within the 750-kb CII *Sna*BI fragment in the Trp⁻ strain.

RESULTS

Screening for auxotrophs: Using Tn*5* mutagenesis of *R. sphaeroides* 2.4.1ΔS, we recovered 33 auxotrophic strains. Five of these strains, CM01, CM02, CM03, CM05, and CM06, were auxotrophic for tryptophan (Mackenzie *et al.* 1995). Strains CM03 and CM05 were capable of growth on minimal medium; however, they took 7–8 days to form colonies compared to 3–4 days for the wild type. The addition of tryptophan to their media restored their growth to wild-type rates. The other four strains did not show any visible growth unless tryptophan was added to the minimal media.

Placing Tn*5* insertions on the physical map: Digestion of *R. sphaeroides* strain 2.4.1ΔS with the enzyme *Spe*I yielded a CI fragment of 735 kb. Digestion of strains CM01, CM02, CM03, and CM05 with *Spe*I resulted in the loss of this fragment. In its place, two fragments, each ~365 kb in size, were generated (Figure 1A). This indicated that the site of Tn*5* insertion in these strains was on CI, within the central region of the 735-kb *Spe*I fragment. These strains showed resolvable restriction pattern differences, indicating that they were the result of independent transposition events. Further mapping studies (not shown) have localized their position to that shown on the physical map (Figure 1A). Because of their proximity, we have marked their position as a single chromosomal location in the figure.

Digestion of the DNA of strain 2.4.1ΔS with the enzyme *Sna*BI yields a CII fragment of 784 kb (Figure 1B).

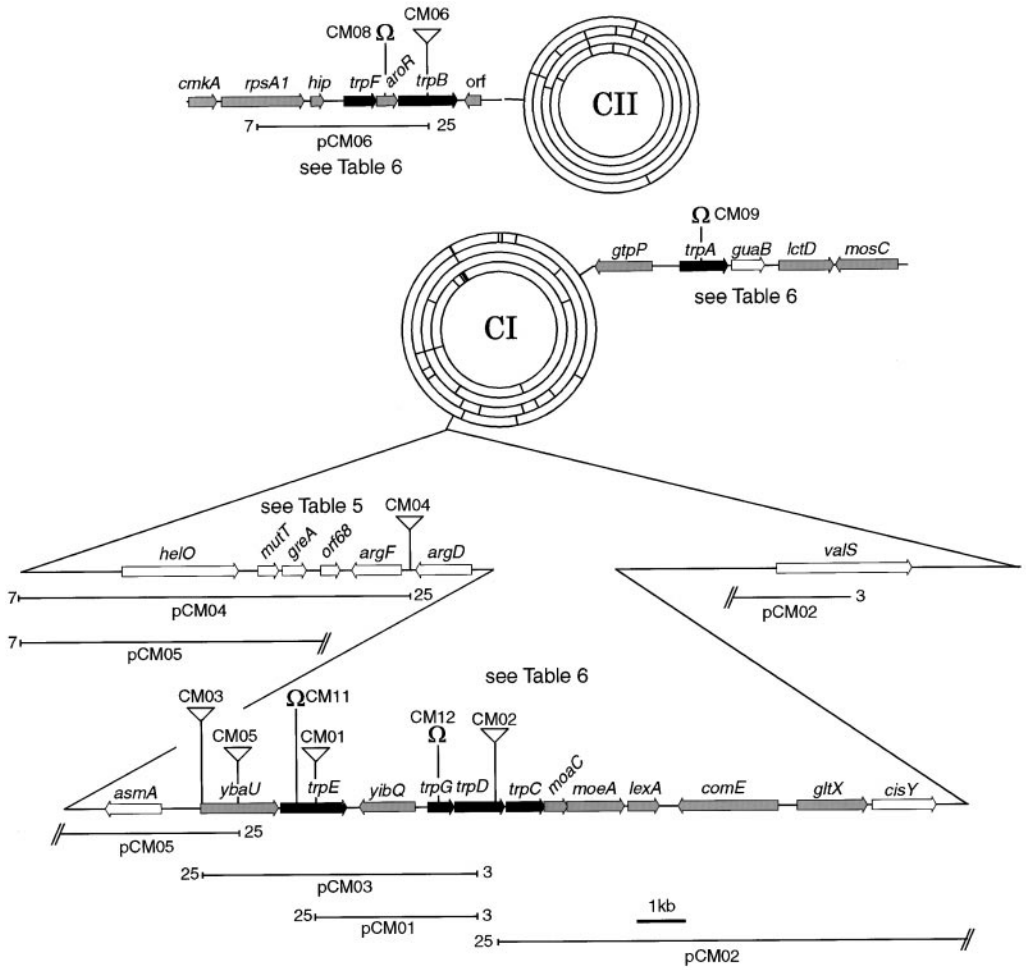

Figure 2.—Overview of gene organization. The three regions encoding the *trp* genes and their flanking sequences have been placed on the physical maps of CI and CII. Genes are shown as arrows indicating the direction of transcription. The *trp* genes are shown as black arrows, fully sequenced genes as gray arrows, and partially sequenced genes as white arrows. The sizes of the partially sequenced genes have been estimated from the size of the genes in the database that they matched. The location of Tn*5* insertions and Ω Sm/Sp cartridge insertions are indicated by inverted triangles and Ω symbols, respectively. The exact locations of these insertions are provided in Table 3. Adjacent to the insertion is the name of the strain in Table 1 that carries that insertion. Beneath the genes are horizontal lines representing the *R. sphaeroides*-Tn*5* *Eco*RI hybrid fragments with the names of the plasmids that carry them (Table 1). At the end of these fragments are the numbers 3, 7, or 25. These indicate the primers T3, T7, and GW25, which were used for sequencing the ends that gave database matches (Table 4). Inserts from plasmids pCM02 and pCM05 are split between the two levels of the CI "Christmas tree." All distances are shown approximately to scale; a scale bar is provided. Note that the Tn5 insertion strain CM04 (on the upper left branch of the tree) is an arginine auxotroph (Arg^-). We looked briefly at this strain, which contained a Tn*5* insertion that mapped 6 kb upstream of *trpE*. The Tn*5*-*R.sphaeroides* hybrid fragment from this strain was subcloned (pCM04) and sequenced using T3, T7, and GW25 primers (Table 4). Additional mapping, subcloning, and sequencing were also carried out. We did not fully sequence this region, but we obtained BLASTX matches (Table 5) and sufficient data to map the region as shown.

Digestion of the DNA of strain CM06 with this enzyme resulted in the loss of this fragment, which was replaced by two restriction fragments of 34 and 750 kb in size. This indicated that the site of this insertion was on CII. Further mapping studies (not shown) placed the Tn*5* insertion at the position shown on the physical map (Figure 1B).

Cloning and sequencing the DNA flanking the site of Tn*5* insertion: We obtained Tp^r *Eco*RI subclones used for cloning and sequencing of each of the auxotrophic strains and used the strategy described in materials and methods. The use of the sequencing primer GW25 revealed that in CI insertion strains CM01 and CM02, the transposon was located in the genes *trpE* and *trpD* (Table 4, Figure 2, and Table 5), which encode the enzymes anthranilate synthase (component I) and anthranilate phosphoribosyltransferase, respectively. The use of primer GW25 to sequence pCM06 revealed that in CII insertion strain CM06, the transposon was located within *trpB* (Figure 2 and Table 4), which encodes the β-subunit of tryptophan synthase. These results also suggested that in *R. sphaeroides*, the genes for tryptophan biosynthesis were distributed between chromosomes CI and CII. The use of primer GW25 indicated that CI insertion strains CM03 and CM05 had Tn*5* insertions in the gene *ybaU*, and that their insertions lay in opposite orientations within the gene (Figure 2 and Table 4). In *E. coli*, this gene encodes a peptidyl-prolyl *cis-trans*-isomerase, which assists in protein folding. It had been noted that both strains CM03 and CM05 grew very slowly in the absence of tryptophan. This result suggested that transposon insertions in the gene *ybaU* may be polar on the nearby *trp* genes. The possibility that YbaU is a regulator of *trp* gene expression or is required for the folding of their gene products has not been excluded.

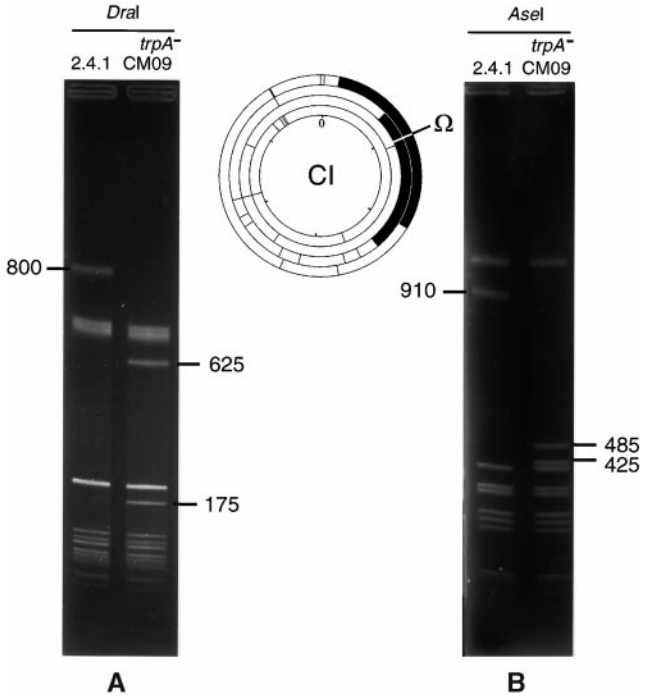

Figure 3.—Mapping *trpA* on CI. A disruption in the *trpA* gene was made using an Ω Sm/Sp cartridge in an *R. sphaeroides* 2.4.1 background. This gave TrpA⁻ strain CM09. Both gel panels show a TAFE gel containing DNA from *R. sphaeroides* 2.4.1 in the left-hand lane. The right-hand lanes contain DNA from strain CM09. Between the gel panels is the physical map of CI, which is composed of four concentric circles marked with restriction sites (moving from outer to inner concentric) for enzymes *Ase*I, *Dra*I, *Sna*BI, and *Spe*I. The zero position (0) is shown at 12 o'clock within the map. The distance markers shown around the inner circle are in increments of 500 kb. The black arcs represent the restriction fragments that are being examined in the gel. The two full arcs are what is found in strain 2.4.1. The four smaller arcs (resulting from cleavage of the full arc caused by Ω Sm/Sp insertion) represent the fragments found in the *trpA*⁻ strain. The site of Ω Sm/Sp is shown by an Ω symbol on the map. The sizes of the fragments described below were determined on additional gels that are not shown. (A) Digestion of 2.4.1 DNA with the restriction enzyme *Dra*I gave a gel band of 800 kb. In the TrpA⁻ strain, this band is absent and is replaced by two smaller bands of 175 and 625 kb. This indicates that the Ω Sm/Sp insertion lies within the 800-kb *Dra*I fragment. (B) Digestion of 2.4.1 DNA with the restriction enzyme *Ase*I gives a gel band of 910 kb. In the *trpA* mutation strain, this band is absent and is replaced by two smaller bands of 485 and 425 kb. This indicates that the Ω Sm/Sp insertion lies near the middle of the the 910-kb *Ase*I fragment. The only position of insertion that will satisfy the results shown in A and B is that shown on the physical map.

Sequencing *trp* genes and surrounding regions on CI: To complete the sequence of *trpD* and trp*E* on CI and to determine if other *trp* genes lay nearby, we sequenced further up- and downstream from the sites of transposon insertion. DNA sequencing also located the precise site of the transposon insertions within the genes (see Table 3). Templates were obtained by subcloning from the *R. sphaeroides*-Tn*5* hybrid *Eco*RI fragments. A DNA region of 14,548 bp was sequenced, and within it were found the *trp* genes, *trpC* and *trpG*, which encode the enzymes indole-3-glycerol phosphate synthase and anthranilate synthase (component II), respectively. This region contained 13 genes, and 11 of these (including *trpE*, *G*, *D*, and *C*) were sequenced to completion (Table 6). BLASTP searching of the database showed that the predicted translations of *trpE*, *G*, *D*, and *C* had high sequence identity to their counterparts in other organisms. The physical map positions of these genes are shown in Figure 2.

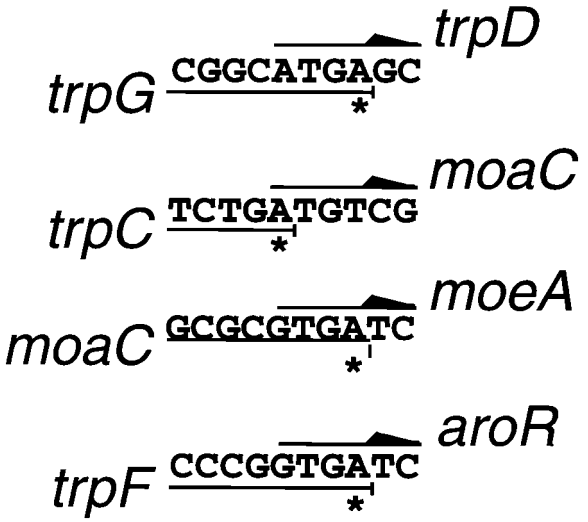

Figure 4.—Overlapping start and stop codons. The overlapping start and stop codons are shown centered on the second base (T) of the ATG or GTG start codons. The A bases of the TGA stop codons are marked with asterisks.

The DNA region between *trpE* and *trpG* did not give a clear match to any database sequence (Figure 2). Computer predictions suggested that a hypothetical gene within this region would encode a protein of 266 amino acid residues. We have called this hypothetical gene *yibQ*, as its predicted translation showed the closest relevant match to YibQ, a hypothetical protein from *Haemophilus influenzae*. This ORF is encoded on the opposite DNA strand to the *trp* genes, raising the possibility that *trpE* and *trpG* may have their own promoters.

In addition to genes for tryptophan biosynthesis, sequencing of this region revealed a number of other genes (Figure 2). Immediately downstream of *trpC* were two putative genes, *moaC* and *moeA*. It has been suggested that MoeA activates molybdenum by conversion to thiomolybdenum (Hasona *et al.* 1998). Both *moaC* and *moeA* are required for molybdopterin synthesis, an essential component of Mo-cofactor-containing enzymes, such as nitrate reductase. In the nodulating bacterium *Bradyrhizobium japonicum*, an operon having the same gene order (*trpD*, *trpC*, and *moaC*-like gene) was found to be essential for plant symbiosis (Kuykendall and Hunter 1997).

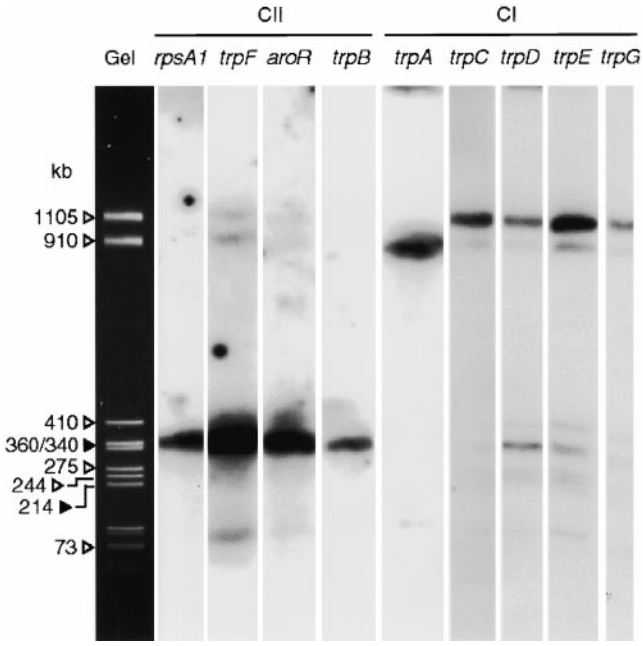

Figure 5.—Localization of genes to the physical map by hybridization. *R. sphaeroides* 2.4.1 DNA was digested with *Ase*I. Ten lanes of a TAFE gel were loaded with equivalent amounts (one-eighth of a plug) of DNA. The gel was run and then photographed. The panel marked "Gel" is representative of the other lanes. The sizes of the *Ase*I fragments (in kilobases) are shown on the left-hand side of the Gel strip. The gel bands are marked with arrows; white and black arrows indicate CI and CII bands, respectively. Bands without arrows are derived from other endogenous replicons. The gel was blotted, and the filter was cut into strips. The strips were then probed with radiolabeled PCR products generated from the genes named at the top of the strips. The autoradiograph strips have been realigned with the Gel strip. Probes to the genes, *rpsA1*, *trpF*, *aroR*, and *trpB* showed strong hybridization to the 360-kb CII *Ase*I band. The *trpA* probe showed strong hybridization to the 910-kb CI *Ase*I band. Probes to the genes, *trpC*, *trpD*, *trpE*, and *trpG* showed strong hybridization to the 1105-kb CI *Ase*I fragment. These data are in agreement with the findings of Tn*5* mapping.

Downstream of these genes lay a putative *lexA*/*dinR* gene. In *E. coli*, *lexA* encodes a repressor of the SOS genes, such as *recA* and *uvrABC* (DNA damage repair genes). BLASTP analysis indicated that the *R. sphaeroides* gene product showed greater homology to proteins from Gram-positive (DinR) than Gram-negative (LexA) bacteria. Previous work suggested that a RecA$^-$ strain of *R. sphaeroides* was less sensitive to ultraviolet light than a RecA$^-$ *E. coli* strain. The differences in sensitivity could not be explained in terms of G + C% composition or target size (Calero *et al.* 1994; Mackenzie *et al.* 1995). This new result adds to the possibility that the DNA repair mechanisms in *R. sphaeroides* could be regulated more like those in Gram-positive bacteria, such as *Bacillus subtilis*.

DNA sequences downstream of *lexA* showed matches to glutamyl-tRNA synthetases (encoded by *gltX*) and citrate synthases (*cisY* partial gene), the latter being a key enzyme in the TCA cycle. The region between *lexA* and *gltX* (and encoded by the opposite DNA strand) gave matches to several ComE proteins. These proteins, encoded by *comE* genes, are involved in competence and DNA uptake in Gram-positive bacteria; however, there is no evidence to suggest that *R. sphaeroides* is naturally competent.

Introducing Ω Sm/Sp insertions into *trpE* and *trpG*: To test the hypothesis that *trpG* was a functional gene, *trpG* was disrupted using an Ω Sm/Sp cartridge (CM12). As a control, we disrupted *trpE* in the same way (CM11). Insertions in *trpE* and *trpG*, which were confirmed by Southern blot analysis (not shown), conferred an auxotrophic phenotype (Trp$^-$). However, the possibility exists that both *trpG* and *trpE* are nonfunctional, and that insertions in these genes have a polar effect on the downstream functional genes *trpDC*. Given their high sequence homology to other genes in the data base, we consider this possibility unlikely. Southern hybridization data (discussed below) further quell this conclusion.

Sequencing *trp* genes and surrounding regions on CII: The DNA region on CII-neighboring *trpB* was sequenced using subclones generated from cosmid pUI8536. This cosmid formed part of an ordered set of clones defining CII, and had been mapped independently to the CII site of the Tn*5* insertion. The Trp$^-$ CII insertion strain, CM06, was restored to prototrophy by complementation when this cosmid was introduced by mating (CM07). When the cosmid vector (pLA2917) was introduced into CM06, it remained Trp$^-$. This provided additional evidence that the Trp$^-$ phenotype was the result of the Tn*5* insertion on CII and not the result of a second mutation at a different chromosomal location.

A CII region of 6203 bp from cosmid pUI8536 was sequenced, and the DNA was found to encode in the order *cmkA*, *rpsA1*, *hip*, *trpF*, *aroR*, and *trpB* (Figure 2 and Table 6). The predicted translation of *trpB* indicated that it encoded 409 amino acid residues, and in strain CM06, the transposon insertion was at codon 91. This suggested that the Tn*5* insertion in *trpB* on CII had resulted in a Trp$^-$ phenotype. Downstream of *trpB* and in the opposite orientation, a region matching a Synechocystis open reading frame (ORF) of unknown function was found.

The CII region upstream of *trpF* encoded three additional genes. The most distal of these, *cmkA* (formerly *mssA*), encodes cytidine monophosphate kinase. In *E. coli*, this gene is considered essential and is required to maintain the normal rate of chromosomal replication. Downstream of *cmkA*, we found the gene *rpsA1*. In *E. coli*, this gene is essential for translation and encodes the largest protein component of the ribosome (Kitakawa and Isono 1982). It has also been shown (in *E. coli*) that *cmkA* is cotranscribed with *rpsA1* as part of the *rpsA* operon (Fricke *et al.* 1995). The gene *hip* (*himD*) lies

TABLE 4

Database matches obtained with DNA sequences flanking the site of transposon insertion

Parent strain	Subclone	Insert size (kb)[a]	BLASTX matches		
			GW25[b]	T3[b]	T7[b]
CM01	pCM01	3.8 (7.1)	*trpE*	*trpD*	Tn*5*
CM02	pCM02	14.3 (17.4)	*trpD*	*valS*[c]	Tn*5*
CM03	pCM03	6.7 (10.0)	*ybaU*	*trpD*	Tn*5*
CM04	pCM04	7.7 (11.0)	*argF*	Tn*5*	No match
CM05	pCM05	13.2 (16.5)	*ybaU*	Tn*5*	No match
CM06	pCM06	3.7 (7.0)	*trpB*	Tn*5*	*rpsA1*

[a] Two numbers are given for each insert size. The first number is the estimated size of cloned *R. sphaeroides* DNA. The numbers in parentheses are the total sizes of the cloned genomic *Eco*RI fragments. This size is the sum of the *R. sphaeroides* DNA plus 3.3 kb of Tn*5* (Tp^r) DNA, which facilitated cloning.
[b] The primer used for sequencing.
[c] The gene *valS* encodes valyl-tRNA synthetase.

downstream of *rpsA1*. This gene encodes the β-subunit of the integration host factor (IHF, Rice *et al.* 1996). The α-subunit of this protein is encoded by the gene *himA* and has been shown to map to the 1105-kb CI *Ase*I fragment (P. Sen and S. Kaplan, unpublished results). We have partially sequenced the region immediately upstream of *cmkA* (data not shown). The result suggested that the gene *aroA*, which encodes the enzyme 3-phosphoshikimate-1-carboxyvinyltransferase, is within this region. A similar gene organization (*aroA-ycaL-cmkA-rpsA1-himD*) has been found in the region 960217–964217 of the *E. coli* genome. However, in *E. coli*, this region is neither followed nor preceded by the genes for tryptophan biosynthesis.

The region between *trpF* and *trpB* was used in a BLASTX search of the database. The result suggested that this region encoded a regulator of the C-P lyase pathway. An Ω Sm/Sp insertion within this gene (CM08) did not result in a Trp$^-$ phenotype. This suggested that *trpB* is not transcribed from the *trpF* promoter. Rather, it has its own promoter and lies downstream of the Ω insertion, *i.e.*, within the 323-bp region preceding the *trpB* start codon. The function of the gene lying between *trpF* and *trpB* is unknown. We have named it *aroR* (*aro*matic amino acid *r*egulator) to reflect its location and a plausible function.

The isolation, sequencing, mapping, and complementation of *trpA* on CI: It had been expected that *trpA* would be downstream of *trpB*, as this is the gene organization in every member of the α-3 group of Proteobact-

TABLE 5

BLASTX matches in the *argDF* region

Putative gene	Gene product/function	Organism match[a]	Polypeptide size[b]	Amino acid identity[c] (%)	Amino acid positives[d] (%)	Expect value[e] (E)
argD	*N*-Acetylornithine aminotransferase	*Methanococcus jannaschii*	398	172/377 (45)	230/377 (60)	3e-85
argF	Ornithine carbamoyltransferase	*Archaeoglobus fulgidus*	307	140/307 (45)	192/307 (61)	4e-71
orf68	Unknown	*Methylobacterium extorquens*	68	26/60 (43)	34/60 (56)	5e-04
greA	Transcription elongation factor	*E. coli*	153	31/92 (33)	50/92 (53)	4e-04
mutT	Mismatch repair	*Methanobacterium thermoautotrophicum*	135	35/97 (36)	49/97 (50)	1e-04
helO	RNA helicase	*Sinorhizobium meliloti*	821	162/405 (40)	225/405 (55)	3e-67

See also Figure 2.
[a] Organism showing the best match to the *R. sphaeroides* sequence.
[b] The size of the matching polypeptide (amino acid residues) encoded by the gene from the organism showing the best match.
[c] The level of amino acid identity between the *R. sphaeroides* sequence and the best match in the database; *i.e.*, for *argD*, the *R. sphaeroides* sequence aligned with 377 amino acid residues of the best match. Of these residues, 172 were identical. This gives an identity of 45%.
[d] Same as for the previous footnote, but including identical plus similar residues.
[e] The probability of getting a match with a similar score by chance.

TABLE 6

BLASTP matches from the CI and CII *trp* regions

Gene	Protein name / predicted function	Matching species	Codons[a]	% Identical[b]	% Positives[c]	E value[d]
		CI *trpE-yibQ-trpDCG* region				
asmA[e]	Protein assembly	*E. coli*	150	23	47	8-04
ybaU	Protease maturation protein	*E. coli*	621	24	42	1e-30
trpE[f]	Anthranilate synthase (Component I)	*R. sphaeroides*	503	99	99	0.0
yibQ	Hypothetical HI0755	*H. influenzae*	266	22	39	0.094
trpG	Anthranilate synthase (Component II)	*Azospirillum brasilense*	195	71	79	2e-71
trpD	Anthranilate phosphoribosyltransferase	*Bradyrhizobium japonicum*	338	57	72	2e-106
trpC	Indoleglycerolphospate synthase	*A. brasilense*	270	64	75	1e-87
moaC	Molybdopterin biosynthesis	*Br. japonicum*	159	62	75	8e-50
moeA	Molybdopterin biosynthesis	*E. coli*	394	34	50	8e-54
lexA/dinR	SOS response regulator	*B. subtilis*	228	34	56	9e-30
comE	Competence protein	*B. subtilis*	729	26	44	7e-27
gltX	Glutamyl-tRNA synthetase	*Rickettsia prowazekii*	471	43	64	e-116
cisY[e]	Citrate synthase	*Br. japonicum*	42	56	73	3e-05
		CII *trpF-aroR-trpB* region				
cmkA	Cytidylate kinase 1	*H. influenzae*	206	41	60	5e-38
rpsA1	Ribosomal protein S1	*Rhizobium meliloti*	561	65	77	0.0
hip	Integration host factor β-subunit	*R. capsulatus*	93	83	90	7e-38
trpF	Phosphoribosylanthranilate isomerase	*Caulobacter crescentus*	212	46	61	1e-45
aroR	Regulator	*Streptomyces griseus*		48	64	8e-31
trpB	Tryptophan synthase β chain	*Pseudomonas aeruginosa*	409	73	83	e-173
orf	Unknown	*Synechocystis* sp.	124	57	69	7e-33
		CI *trpA* region				
gtpP	GTP binding protein	*B. subtilis*	369	55	72	e-115
trpA	Tryptophan synthase α chain	*P. putida*	263	58	73	8e-86
guaB[e]	Inosine-5′-monophosphate dehydrogenase	*Methanobacterium thermoautotrophicum*	71	38	63	1e-06
lctD	Lactate dehydrogenase	*Neisseria meningitidis*	387	65	82	e-143
mosC	Transporter	*Rhizobium* sp. NGR234	385	25	42	6e-28

[a] The number of codons (amino acid residues) that encode the *R. sphaeroides* gene (excluding the stop codon).

[b,c] Percentage of amino acid identity and similarity between the predicted translation of the *R. sphaeroides* protein and the protein in the database that showed the best BLASTP match.

[d] Expect value given from an unfiltered BLASTP search.

[e] This gene has not been fully sequenced.

[f] The second best match was to *P. putida* TrpE (493 codons). Identity, 51%; Positive, 64%; Expect, e-125.

eria examined to date. As a result, we carried out PCR using *R. sphaeroides* genomic DNA as a template and primers designed to *trpA* of *R. capsulatus*. A 600-bp DNA product was generated. After DNA sequencing, it gave a BLASTX match to database TrpA proteins, which are the α-subunits of tryptophan synthase. This PCR product was used as a probe to screen an *R. sphaeroides* cosmid library, to which five cosmid clones hybridized. Their DNA was purified, and they were probed in a Southern blot with the *trpA* PCR product (result not shown). The result showed that *trpA* was located on a 6.5-kb *Bam*HI fragment. This fragment was subcloned from cosmid pUI8063, and the region sequenced. Within this region, a putative *trpA* gene and four other putative genes were found (see Figure 2 and Table 6).

An Ω Sm/Sp cartridge was inserted into the cloned *trpA* gene (pCM10A). This gene interruption was introduced into the *R. sphaeroides* genome as described in materials and methods. From five independent matings, 34 Sm^r/Sp^r strains were isolated. Five of these were found to be Tc^s. These 5 strains were also tryptophan auxotrophs (Trp^-), suggesting that in these strains, a

chromosomal interruption of *trpA* had occurred by a double-crossover event. This hypothesis was validated by genomic Southern blot (not shown). One of these TrpA$^-$ strains was designated CM09.

To further confirm that the observed Trp$^-$ phenotype was a result of the disruption of the *trpA* gene, cosmid pUI8063 (*trpA*$^+$ Tc^r) or the cosmid vector pLA2917 (Tc^r) was introduced by mating into each of the five *R. sphaeroides* TrpA$^-$ strains. Twenty colonies from each mating were restreaked on SMM Tc and SMM Tc tryptophan plates. Colonies that had received the vector alone grew only on SMM Tc with tryptophan. Colonies that had received cosmid pUI8063 grew on all three plates (*e.g.*, CM10). This suggested that the disruption of the *trpA* gene in the five mutants was responsible for the Trp$^-$ phenotype.

The Ω Sm/Sp cartridge carries recognition sites for the restriction enzymes *Ase*I and *Dra*I. Use of these sites allowed us to locate *trpA* on the physical map. We mapped the five Tc^s TrpA$^-$ strains to the same location on the map by TAFE pulse-field gel electrophoresis. The mapping of one of these strains, CM09, is shown as an example in Figure 3. It can be seen that in the 2.4.1 wild-type strain *Dra*I (Figure 3A) and *Ase*I (Figure 3B), digestion generates CI fragments of 800 and 910 kb, respectively. In the TrpA$^-$ mutant, these fragments are absent. They are replaced by two CI *Dra*I fragments of 625 and 175 kb (Figure 3A) and two CI *Ase*I fragments of 485 and 425 kb (Figure 3B). These results suggested that *trpA* and *trpB* were on different chromosomes, and that *trpA* was located on the physical map ~1.23 Mb (149° counterclockwise) from the other CI *trp* genes (Figures 2 and 3).

Gene fusions overlapping stop-start codons and ribosome-binding sites: It had been shown previously that a number of genes in the tryptophan pathway are fused. For example, in *Rhz. meliloti* and its relatives, *trpE* and *trpG* are fused to give *trp*(*EG*), resulting in a fusion of the α- and β-subunits of anthranilate synthase into a single polypeptide. Examination of the *trp* genes of *R. sphaeroides* indicated that such gene fusions had not evolved. Of note, however, were the numbers of neighboring genes that shared overlapping stop and start codons (Figure 4); *i.e.*, *trpG* stop overlaps with *trpD* start, *trpC* stop overlaps with *moaC* start, *moaC* stop overlaps with *moeA* start, and *trpF* stop overlaps with *aroR* start. The genes *trpC*, *moeA*, and *aroR* had putative ribosome-binding sites upstream of their start codons; however, ribosome-binding sites were not found upstream of the genes *trpG*, *trpD*, and *moaC*. All other tryptophan-related genes, *trpA*, *trpB*, *trpE*, and *ybaU*, were found to have putative ribosome-binding sites upstream of their initiation codons.

TAFE Southern blot hybridization with *trp*, *rpsA1*, and *aroR* probes: Internal primer pairs (Table 2) were made to the following genes: *rpsA1*, *trpF*, *aroR*, *trpB*, *trpA*, *trpC*, *trpD*, *trpE*, and *trpG*, and were used for PCR. After sequencing had verified that the expected fragments had been generated, they were used to probe *R. sphaeroides* *Ase*I TAFE Southern blots (Figure 5). The results suggested that *trpA* was located on CI within the 910-kb *Ase*I fragment, and that *trpC*, *trpD*, *trpE*, and *trpG* were located on CI within the 1105-kb *Ase*I fragment. Genes *rpsA1*, *trpF*, *aroR*, and *trpB* were found to be located on CII. A shorter exposure (not shown) indicated that these genes hybridized to the 360-kb CII *Ase*I fragment rather than to the slightly smaller 340-kb CII *Ase*I fragment.

It was noted that some of the probes hybridized less strongly, but visibly, to other *Ase*I fragments. To determine if there were additional silent copies of these genes, we used them to probe regular (non-TAFE) *Bam*HI and *Eco*RI genomic Southern blots (not shown). The results firmly indicated that in the TAFE Southern blots we were observing nonspecific hybridization to abundant, large TAFE fragments. In standard Southern blots, we could detect only single copies of all the genes described.

DISCUSSION

Transposon mutagenesis was used to generate *R. sphaeroides* auxotrophs with a Trp$^-$ phenotype. The sites of Tn*5* insertion were determined by TAFE gel electrophoresis and were mapped to CI and CII. These results suggested that the genes encoding the tryptophan biosynthetic pathway were distributed between the two chromosomes of this multichromosomal bacterium. Sequencing of the regions around the sites of Tn*5* insertion indicated that transposons had disrupted the CI genes *trpE*, *trpD*, and *ybaU*, as well as the CII gene *trpB*. The insertions in *ybaU* are thought to have resulted in a Trp$^-$ phenotype because of polar effects on the downstream gene *trpE*. Additional sequencing revealed the genes *trpG* and *trpC* on CI and *trpF* on CII. Additional cloning indicated that *trpA* was located on the CI physical map 1.23 Mb (149° counterclockwise) from the other CI *trp* genes. This accounted for all the structural genes of the classical tryptophan pathway.

To further verify function, the genes *trpA*, *trpE*, and *trpG* were disrupted with an Ω Sm/Sp cartridge. In each case the disruption led to a Trp$^-$ phenotype, confirming the role predicted from sequence analysis. The disruption of *trpE* for a second time (the first being with Tn*5*) acted as a control for Tn*5* mutagenesis. This result suggested that the Trp$^-$ phenotypes were the result of the Tn*5* insertions and had not arisen because of a mutation at a second chromosomal location. This was further confirmed by complementation of the TrpA$^-$ and TrpB$^-$ strains with cosmids carrying the wild-type genes. A mutation at a second location was unlikely to have been complemented by these cosmids. In addition, the finding that disruption of these genes resulted in auxotrophy suggested that these genes are found in single copy within the genome. Southern hybridization confirmed this finding and corroborated the location of the mapped insertions. This has led us to conclude that

the structural genes corresponding to the tryptophan biosynthetic pathway are indeed distributed between the two chromosomes of *R. sphaeroides* 2.4.1.

A number of the *trp* genes have overlapping stop and start codons. Such an organization has been noted previously in other organisms; *e.g.*, in *E. coli*, *trpB* and *trpA* coding regions on the polycistronic *trp* mRNA are separated by overlapping stop and start codons. Efficient translation of the *trpA* coding region is subject to translational coupling; *i.e.*, maximal *trpA* expression is dependent on prior translation of the *trpB* coding region. Therefore, it is both possible and understandable that the gene pairs *trpG-trpD* and *trpF-aroR* are translationally coupled. However, it is less obvious why translational coupling (if it does indeed occur) would be present between *trpC-moaC-moeA*. These last two genes have been implicated in molybdopterin biosynthesis. We have found a second copy of *moeA* on CII, downstream of *torA*, a gene that encodes the molybdenum-containing enzyme TMAO reductase (N. Mouncey and S. Kaplan, unpublished results).

In addition to *trpF*, *aroR*, and *trpB*, other genes, *i.e.*, *cmkA*, *rpsA1*, and *hip*, were also mapped to CII. Southern hybridization suggested that *rpsA1* is located only on this chromosome. Their low BLASTP scores, combined with the finding that they are in a similar genomic order as in other bacteria, suggests that these are *bona fide* genes. This further reinforces previous work that suggested that CII encodes functions typical of any other bacterial chromosome.

However, bacteria that possess multiple chromosomes are an enigma. What selects for the maintenance of a divided genome once such an event has occurred? The possession of two chromosomes surely increases the complexity of cell division, imparting an increased risk for genetic lesions and decreased fitness of daughter cells. This is particularly noticeable in *R. sphaeroides*, where partial loss of CII could lead to auxotrophy, or inability to carry out translation or genome replication.

Multiple chromosomes would also be expected to lead to an increase in the complexity of coordinated gene expression, as supported by current reasoning. For example, if we examine the enzyme tryptophan synthase, we see that in most bacteria it is encoded by a *trpBA* operon. This "makes sense" because the cell needs to make equimolar amounts of the α- and β-subunits to form the functional heterotetrameric ($\alpha_2\beta_2$) enzyme. In these bacteria, transcription of both genes is coordinated, and translation products are synthesized near each other to form the complete enzyme. This would appear to be an efficient and highly evolved process. However, in *R. sphaeroides*, the genes that encode these subunits are on different chromosomes. How does the cell ensure that the gene products are formed in equimolar amounts and in relative proximity for subunit association? Perhaps it does not, at least not to the same degree as in other bacteria. It may be that making different amounts of the two products does not decrease the fitness of the cell sufficiently for there to have been a strong selective drive toward operon formation. Therefore, in *R. sphaeroides*, the *trp* genes may be organized in a more ancient and perhaps less stringently regulated topology than normally seen in bacteria. Reinforcing this hypothesis is the finding that in *Aquifex aeolicus*, a member of the deepest branching family within the bacterial domain, the *trp* genes are distributed as individual genes throughout the genome (Deckert *et al.* 1998). Indeed, *A. aeolicus* is extreme in that no two amino acid biosynthetic genes are found in the same operon. If the "disbursed genes are ancient and operons are recent" argument is accepted, then by analogy, having multiple chromosomes may also represent a more ancient and less "streamlined" genomic organization. Multiple chromosomes may persist, not because such an organization confers a biological advantage, but rather because it does not confer a sufficient decrease in fitness to have been selected against and lost from the genome pool.

Other members of the α-Proteobacteria, *e.g.*, Agrobacteria and Brucella, also have genomes comprising multiple chromosomes (or, as in the case of Rhizobia, megaplasmids), and it has been noted that many of these have infectious associations with eukaryotes. Evidence from 16S ribosomal RNA suggests that ancient members of this group gave rise to eukaryotic plastids, such as the chloroplast and mitochondrion (Gray 1993). Therefore, complex genomic organization with "disbursed" as opposed to "operonic" or "condensed" gene organization may have been a prerequisite for the formation of early eukaryotic cells.

In *R. sphaeroides*, it is known that extensive gene duplication occurs between the two chromosomes. If a Rhodobacter-like organism was the progenitor of such plastids, then gene duplication could have provided the opportunity for the development of complex regulatory mechanisms, such as those found in eukaryotes. Such duplications may have permitted the evolution of differential regulation of each copy of the duplicated gene, resulting in a wider spectrum of conditions under which a gene or group of genes with similar function could be expressed.

The possession of multiple chromosomes also may have assisted in genetic export and exchange between the early plastids and their host, perhaps explaining why many plastid genes are encoded in the nucleus of modern eukaryotes. This suggests that if we examined the plastids of primitive unicellular eukaryotes, we may see remnants of their ancient bacterial origins. It is therefore intriguing that in the "primitive" unicellular red alga *Cyanidium caldarium*, the *trpA* gene is found on the plastid genome. In contrast, *trpB* is located in the cell nucleus (Ohta *et al.* 1994). Thus, the very unique structure of the *R. sphaeroides* genome, the dispersal of essential genetic loci between linkage groups, and the organism's unique position within the α-Proteobacteria

suggest its centrality to the origins of primitive eubacteria and plastids.

We are grateful to Agnes Puskas, Renata Ng, and David Needleman for their help with DNA sequencing; Mark Gomelsky for technical tips; and Jesus Eraso for technical tips and correcting our manuscript. We are also grateful to Robert Hazelkorn for providing us with the *R. capsulatus trpA* subclones and sequence before their publication. We also thank our reviewers. Their constructive criticism led to a greatly improved manuscript. This work was supported by National Institutes of Health grant GM-55481.

LITERATURE CITED

Allardet-Servent, A., S. Michaux-Charachon, E. Jumas-Bilak, L. Karayan and M. Ramuz, 1993 Presence of one linear and one circular chromosome in the *Agrobacterium tumefaciens* C58 genome. J. Bacteriol. **175:** 7869–7874.

Allen, L. N., and R. S. Hanson, 1985 Construction of broad-host range cosmid cloning vectors: identification of genes necessary for growth of *Methylobacterium organophilum* on methanol. J. Bacteriol. **161:** 955–962.

Baril, C., J. L. Herrmann, C. Richaud, D. Margarita and I. S. Girons, 1992 Scattering of the rRNA genes on the physical map of the circular chromosome of *Leptospira interrogans* serovar icterohaemorrhagiae. J. Bacteriol. **174:** 7566–7571.

Calero, S., A. R. Fernandez De Henestrosa and J. Barbe, 1994 Molecular cloning, sequence and regulation of expression of the *recA* gene of the phototrophic bacterium *Rhodobacter sphaeroides*. Mol. Gen. Genet. **242:** 116–120.

Choudhary, M., C. Mackenzie, K. S. Nereng, E. J. Sodergren, G. Weinstock *et al.*, 1994 Multiple chromosomes in bacteria: structure and functions of chromosome II of *Rhodobacter sphaeroides* 2.4.1^T. J. Bacteriol. **176:** 7694–7702.

Choudhary, M., C. Mackenzie, K. Nereng, E. Sodergren, G. M. Weinstock *et al.*, 1997 Low-resolution sequencing of *Rhodobacter sphaeroides* 2.4.1^T: chromosome II is a true chromosome. Microbiology **143:** 3085–3099.

Choudhary, M., C. Mackenzie, N. J. Mouncey and S. Kaplan, 1999 RsGDB, the *Rhodobacter sphaeroides* genome database. Nucleic Acids Res. **27:** 61–62.

Crawford, I. P., 1989 Evolution of a biosynthetic pathway: the tryptophan paradigm. Annu. Rev. Microbiol. **43:** 567–600.

Deckert, G., P. V. Warren, T. Gaasterland, W. G. Young, A. L. Lenox *et al.*, 1998 The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature **392:** 353–358.

Donohue, T. J., and S. Kaplan, 1992 Genetic techniques in *Rhodospirillaceae*. Methods Enzymol. **204:** 459–485.

Dryden, S., and S. Kaplan, 1990 Localization and structural analysis of the ribosomal RNA operons of *Rhodobacter sphaeroides*. Nucleic Acids Res. **18:** 7267–7277.

Fricke, J., J. Neuhard, R. A. Kelln and S. Pedersen, 1995 The *cmk* gene encoding cytidine monophosphate kinase is located in the *rpsA* operon and is required for normal replication rate in *Escherichia coli*. J. Bacteriol. **177:** 517–523.

Gray, M. W., 1993 Origin and evolution of organelle genomes. Curr. Opin. Genet. Dev. **3:** 884–890.

Hallenbeck, P. L., R. Lerchen, P. Hessler and S. Kaplan, 1990a Phosphoribulokinase activity and the regulation of CO_2 fixation critical for photosynthetic growth of *Rhodobacter sphaeroides*. J. Bacteriol. **172:** 1749–1761.

Hallenbeck, P. L., R. Lerchen, P. Hessler and S. Kaplan, 1990b The role of CFXA, CFXB, and external electron acceptors in the regulation of ribulose 1,5-bisphosphate carboxylase/oxygenase expression in *Rhodobacter sphaeroides*. J. Bacteriol. **172:** 1736–1748.

Hasona, A., R. M. Ray and K. T. Shanmugam, 1998 Physiological and genetic analyses leading to identification of a biochemical role for the *moeA* (molybdate metabolism) gene product in *Escherichia coli*. J. Bacteriol. **180:** 1466–1472.

Heath, J. D., J. D. Perkins, B. Sharma and G. M. Weinstock, 1992 *Not*I genomic cleavage map of *Escherichia coli* K-12 strain MG1655. J. Bacteriol. **174:** 558–567.

Jumas-Bilak, E., S. Michaux-Charachon, G. Bourg, M. Ramuz and A. Allardet-Servent, 1998 Unconventional genomic organization in the alpha subgroup of the Proteobacteria. J. Bacteriol. **180:** 2749–2755.

Kishan, V., and W. Hillen, 1990 Molecular cloning, nucleotide sequence, and promoter structure of the *Acinetobacter calcoaceticus trpFB* operon. J. Bacteriol. **172:** 6151–6155.

Kitakawa, M., and K. Isono, 1982 An amber mutation in the gene *rpsA* for ribosomal protein S1 in *Escherichia coli*. Mol. Gen. Genet. **185:** 445–447.

Kuykendall, L. D., and W. J. Hunter, 1997 The sequence of a symbiotically essential *Bradyrhizobium japonicum* operon consisting of *trpD*, *trpC* and a *moaC*-like gene. Biochim. Biophys. Acta **1350:** 277–281.

Mackenzie, C., M. Chidambaram, E. J. Sodergren, S. Kaplan and G. Weinstock, 1995 DNA repair mutants of *Rhodobacter sphaeroides*. J. Bacteriol. **177:** 3027–3035.

Michaux, S., J. Paillisson, M.-J. Carles-Nurit, G. Bourg, A. Allardet-Servent *et al.*, 1993 Presence of two independent chromsomes in the *Brucella melitensis* 16M genome. J. Bacteriol. **175:** 701–705.

Neidle, E., and S. Kaplan, 1992 *Rhodobacter sphaeroides rdxA*, a homolog of *Rhizobium meliloti fixG*, encodes a membrane protein which may bind cytoplasmic [4Fe-4S] clusters. J. Bacteriol. **174:** 6444–6454.

Neidle, E., and S. Kaplan, 1993 Expression of the *Rhodobacter sphaeroides hemA* and *hemT* genes encoding two aminolevulinic acid synthase isozymes. J. Bacteriol. **175:** 2292–2303.

Nichols, B. P., 1996 Evolution of genes and enzymes of tryptophan biosynthesis, pp. 2638–2648 in *Escherichia coli and Salmonella*, edited by F. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low *et al.* American Society for Microbiology, Washington, DC.

Ohta, N., N. Sato, S. Kawano and T. Kuroiwa, 1994 The *trpA* gene on the plastid genome of *Cyanidium caldarium* strain RK-1. Curr. Genet. **25:** 357–361.

Pittard, A. J., 1996 Biosynthesis of the aromatic amino acids, pp. 458–484 in *Escherichia coli and Salmonella*, edited by F. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low *et al.* American Society for Microbiology, Washington, DC.

Rice, P. A., S. Yang, K. Mizuuchi and H. A. Nash, 1996 Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. Cell **87:** 1295–1306.

Rodley, P. D., U. Romling and B. Tummler, 1995 A physical genome map of the *Burkholderia cepacia* type strain. Mol. Microbiol. **17:** 57–67.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Simon, R., U. Priefer and A. Puhler, 1983 A broad host range mobilization system for in vivo genetic engineering: transposon mutagenesis in Gram negative bacteria. Bio/Technology **1:** 37–45.

Suwanto, A., and S. Kaplan, 1989a Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: genome size, fragment identification, and gene localization. J. Bacteriol. **171:** 5840–5849.

Suwanto, A., and S. Kaplan, 1989b Physical and genetic mapping of the *Rhodobacter sphaeroides* 2.4.1 genome: presence of two unique circular chromosomes. J. Bacteriol. **171:** 5850–5859.

Suwanto, A., and S. Kaplan, 1992 Chromosome transfer in *Rhodobacter sphaeroides*: Hfr formation and genetic evidence for two unique circular chromosomes. J. Bacteriol. **174:** 1135–1145.

Tabita, F. R., J. L. Gibson, B. Bowien, L. Dijkhuizen and W. G. Meijer, 1992 Uniform designation for genes of the Calvin-Benson-Bassham reductive pentose phosphate pathway of bacteria. FEMS Microbiol. Lett. **78:** 107–110.

Trucksis, M., J. Michalski, Y. K. Deng and J. B. Kaper, 1998 The *Vibrio cholerae* genome contains two unique circular chromosomes. Proc. Natl. Acad. Sci. USA **95:** 14464–14469.

Winterstein, C., and B. Ludwig, 1998 Genes coding for respiratory complexes map on all three chromosomes of the *Paracoccus denitrificans* genome. Arch. Microbiol. **169:** 275–281.

Woese, C. R., O. Kandler and M. L. Wheeler, 1990 Towards a natural system of organisms: proposal for the domains Archaea, Bacteria and Eucarya. Proc. Natl. Acad. Sci. USA **87:** 4576–4579.

Communicating editor: R. Maurer